



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/896,199	06/29/2001	Eric Cohen-solal	US010324	7568

24737 7590 11/02/2004

PHILIPS INTELLECTUAL PROPERTY & STANDARDS  
P.O. BOX 3001  
BRIARCLIFF MANOR, NY 10510

EXAMINER

BELL, PAUL A

ART UNIT	PAPER NUMBER
----------	--------------

2675

12

DATE MAILED: 11/02/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

**UNITED STATES DEPARTMENT OF COMMERCE****U.S. Patent and Trademark Office**

Address: COMMISSIONER FOR PATENTS

P.O. Box 1450

Alexandria, Virginia 22313-1450

APPLICATION NO./ CONTROL NO.	FILING DATE	FIRST NAMED INVENTOR / PATENT IN REEXAMINATION	ATTORNEY DOCKET NO.
---------------------------------	-------------	---	---------------------

EXAMINER
----------

ART UNIT
----------

PAPER
-------

12

DATE MAILED:

**Please find below and/or attached an Office communication concerning this application or proceeding.**

**Commissioner for Patents**

SEE attached EXAMINER'S ANSWER

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Paul Bell whose telephone number is (703) 306-3019.

If attempts to reach the examiner by telephone are unsuccessful the Technology Center 2600 Customer Service Office whose telephone number is (703) 306-0377 can help with any inquiry of a general nature or relating to the status of this application.

Any response to this action should be mailed to:

Commissioner of Patents and Trademarks  
Washington, D.C. 20231

Or Faxed to: (703) 872-9306

Paul Bell

Art unit 2675

October 29, 2004



UNITED STATES PATENT AND TRADEMARK OFFICE

**RECEIVED**

NOV 02 2004

**Technology Center 2600**

COMMISSIONER FOR PATENTS  
UNITED STATES PATENT AND TRADEMARK OFFICE  
P.O. Box 1450  
ALEXANDRIA, VA 22313-1450  
[www.uspto.gov](http://www.uspto.gov)

**BEFORE THE BOARD OF PATENT APPEALS  
AND INTERFERENCES**

Application Number: 09/896,199  
Filing Date: June 29, 2001  
Appellant(s): COHEN-SOLAL, ERIC

---

Thomas Spinelli  
For Appellant

**EXAMINER'S ANSWER**

This is in response to the appeal brief filed 3/29/2004.

**(1) *Real Party in Interest***

A statement identifying the real party in interest is contained in the brief.

**(2) *Related Appeals and Interferences***

A statement identifying the related appeals and interferences which will directly affect or be directly affected by or have a bearing on the decision in the pending appeal is contained in the brief.

**(3) *Status of Claims***

The statement of the status of the claims contained in the brief is correct.

**(4) *Status of Amendments After Final***

The appellant's statement of the status of amendments after final rejection contained in the brief is correct.

**(5) *Summary of Invention***

The summary of invention contained in the brief is correct.

**(6) *Issues***

The appellant's statement of the issues in the brief is correct.

**(7) *Grouping of Claims***

The rejection of claims 1-20 stand or fall together because appellant's brief does not include a statement that this grouping of claims does not stand or fall together and reasons in support thereof. See 37 CFR 1.192(c)(7).

**(8) *Claims Appealed***

The copy of the appealed claims contained in the Appendix to the brief is correct.

**(9) Prior Art of Record**

5,999,214                                      Inagaki                                      12-1999

Pavlovic et al., "Integration of Audio/Visual Information for Use in Human-Computer Intelligent Interaction", Image Processing, 1997 Proceedings IEEE, pages 121-124.

**(10) Grounds of Rejection**

The following ground(s) of rejection are applicable to the appealed claims:

Claims 1-20 are rejected under 35 U.S.C. 103(a) as being unpatentable over Inagaki (5,999,214) in view of Pavlovic et al. "Integration of audio/visual information for use in human-computer intelligent interaction", Image processing, 1997 Proceedings IEEE pages 121-124.

With regard to claim 1 Inagaki teaches a video display device comprising: a display configured to display a primary image and a picture-in-picture image (PIP) overlaying the primary image (figure 11, items 13 and 17); and a processor operatively coupled to the display and configured to receive a first video data stream for the primary image, to receive a second video data stream for the PIP (figure 11, items 22 and 16),

Inagaki does not teach, "and to change a PIP display characteristic in response to a received audio command and a related gesture from a user". Inagaki apparatus instead detects and responds to any of the many sounds or "audio indications" in the form of a unique voices of a specific speaking attendees with the same command which is move the camera and highlight the PIP of the speaking

Art Unit: 2675

attendee and does not depend on "related gesture from a user" (figure 11 "VOICE DIRECTION DETECTION UNIT", column 3, lines 31-33, column 10, lines 16-25).

However Pavlovic does demonstrate the concept of a system utilizing a combination of "audio commands" and a "related gesture" from a user as a means of controlling a graphical object on display which is analysis to where Inagaki controlled a specific graphical object such as a PIP on a display (SEE Pavlovic page 123 3. EXPERIMENTAL RESULTS section ).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to use a "received audio command and a related gesture from a user", as taught by Pavlovic in the apparatus of Inagaki, because of the motivation directly provided by Pavlovic; "Psychological studies, for example, show that people prefer to use hand gestures in combination with speech in a virtual environment, since they allow the user to interact without special training or special apparatus".

With regard to claim 2 the combination of Inagaki and Pavlovic teaches the video display device of claim 1, wherein the PIP display characteristic is at least one of a position of the PIP on the display and a display size of the PIP (See Inagaki which illustrates at least one of these changes, for example; "a position of the PIP on the display" figure 8a illustrates the concept of a relationship between who is speaking and the position of which PIP to be highlighted which further changes when speaker changes which is clearly illustrated in figure 8B so therefore it reads on this broad language).

With regard to claim 3 the combination of Inagaki and Pavlovic teaches the video display device of claim 1, comprising: a microphone for receiving the audio indication from the user; and a camera for acquiring an image of the user containing the related gesture (See Inagaki figure 11).

With regard to claim 4 the combination of Inagaki and Pavlovic teaches the video display device of claim 1 wherein the processor is configured to analyze audio information received from the user to identify when a PIP related audio indication is intended by the user (See Inagaki figure 8a and 8b).

With regard to claim 5 the combination of Inagaki and Pavlovic teaches the video display device of claim 1, wherein the processor is configured to analyze image information received from the user after the audio indication is received to identify the change in the PIP display characteristic that is expressed by the received gesture (See Inagaki figure 8a and 8b and Pavlovic et al figures 6-8 and especially the Pavlovic figure 5 "HIGH LEVEL FEATURE INTEGRATION" where it was obvious the pre analyze step is to simultaneously receive the video and audio data using the camera and the microphone, where it is then; split into a parallel visual and audio estimator/classifier module which is followed by a second stage which contains a feature integration/combination module where the combination module computes the likelihood of the pairs of gesture and verbal words. This claim language is very broad here because Pavlovic clearly receives both the audio and video before he analyzes the video or audio data, this is just the logical progression claimed).

With regard to claim 6 the combination of Inagaki and Pavlovic teaches the video display device of claim 5, wherein the image information is contained in a sequence of images and wherein the processor is configured to analyze the sequence of images to determine the received gesture (since a gesture can be a motion which would require a sequence of images to detect this feature is obvious to the system of Inagaki and Pavlovic also see Pavlovic section 2.1).

With regard to claim 7 the combination of Inagaki and Pavlovic teaches the video display device of claim 1, wherein the image information is contained in a sequence of images and wherein the processor is configured to determine the received gesture by analyzing the sequence of images and determining a trajectory of a hand of the user (since a gesture can be a motion which would require a sequence of images to detect this feature is obvious to the system of Inagaki and Pavlovic and is merely viewed as directed towards an obvious intended use of which the combination of which it is capable also see Pavlovic section 2.1).

With regard to claim 8 the combination of Inagaki and Pavlovic teaches the video display device of claim 1, wherein the processor is configured to determine the received gesture by analyzing an image of the user and determining a posture of a hand of the user (since a gesture can be a posture of a hand this feature is obvious to the system of Inagaki and Pavlovic and is merely viewed as directed towards an obvious intended use of which the combination of which it is capable also see Pavlovic section 2.1).



With regard to claim 9 the combination of Inagaki and Pavlovic suggest the video display device of claim 1, wherein the video display device is a television (since Pavlovic shows a projection screen in figure 6 and since it is also well-known in the prior art that televisions use projection screens one would be motivated to have a projection screen with a dual use such as conference and watching the game and is merely viewed as directed towards an obvious intended use of which the combination of which it is capable) .

With regard to claim 10 the combination of Inagaki and Pavlovic teaches the video display device of claim 1, wherein the image is a sequence of images of the user containing the user gesture, the video display device comprising a camera for acquiring the sequence of images of the user (see Inagaki figure 11, item 2).

With regard to claims 11-14 most of the limitations was already shown above with regards to apparatus claims 1-10 to be obvious and therefore the method claims 11-14 which corresponds to the apparatus were also obvious and in addition the applicant is now specifically claiming; "determining whether the received audio indication is one of a plurality of expected audio indications: analyzing a gesture of the user if the received audio indication is one of the plurality of expected audio indications" (SEE Pavlovic figure 7 where he illustrates a plurality of "expected audio indications" SPEECH , and a plurality of "expected gestures" GESTURE. Now look at Pavlovic figure 5 where he illustrates in the audio estimator/ classifier module receiving and "determining whether the received audio indication is one of a plurality of expected audio indications" and where also he illustrates in the video estimator/

Art Unit: 2675

classifier module receiving and "determining whether the received gesture is one of a plurality of expected gestures". It is an obvious practice that if either data collection process produces an error because the audio indication or gesture used is not from the expected sets illustrated in figure 7 that the next step of "analyzing a gesture of the user if the received audio indication is one of the plurality of expected audio" in the Feature Integrator will not happen. This is because it is an obvious practice when an artificial intelligent or smart device as illustrated by the combination of Inagaki/Pavlovic can not comprehend the data within a reasonable range of certainly or as stated by Pavlovic "computes the likelihood" that it simply errors out in the flow chart and does nothing but waits for further inputs.)

With regard to claims 15-18 the combination of Inagaki and Pavlovic was shown above to read on most of these limitation in claims 1-14 in addition to summarize a feature directed towards a program stored implementing this process is inherent to the automatic computer system taught by the combination of Inagaki and Pavlovic.

With regard to claims 19 the combination of Inagaki and Pavlovic was shown above to read on all of these limitation in claims 1-18.

With regard to claim 20 the combination of Inagaki and Pavlovic was shown above to read on most of these limitation in claims 1-18 in addition to summarize a specific feature directed towards , "wherein the processor is configured to analyze image information received from the user after the audio indication is received to identify the change in the PIP display characteristic that is expressed by the received gesture" (See Pavlovic figure 5 and specifically the rejection of 11 above).

**(11) Response to Argument**

The applicant argues on pages 10 and 11, with regard to claim 1, that Inagaki does not disclose or suggest, "a processor ...configured...to change a PIP display characteristic in response to a received audio command and a related gesture from a user." And also applicant argues that Independent claims 11, 15, 19 and 20 contain similar recitations regarding an audio command or indication and a related gesture.

It should be noted that only claim 1 has the specific phrase "audio command" all other independent and dependent claims have the phrase "audio indication". It is of interest to note that applicants original specification and claims never uses the phrase "audio command" but instead used the phrase "audio indication" for example section [0009] states, "The system utilizes a combination of an "audio indication" and a related gesture from the user to control PIP display characteristic ". Therefore since the coined phrase "audio command" is not in the original specification it is clearly open to the broadest interpretation reasonably possible and one such interpretation in view of the specification is that it can be interpreted to have the same meaning as "audio indication" as used in the specification.

In response to applicant's arguments against the references individually (Inagaki only), one cannot show nonobviousness by attacking references individually where the rejections are based on combinations of references (Inagaki and Pavlovic). See *In re Keller*, 642 F.2d 413, 208 USPQ 871 (CCPA 1981); *In re Merck & Co.*, 800 F.2d 1091, 231 USPQ 375 (Fed. Cir. 1986).

The examiner disagrees with applicants argument because since the Inagaki system detects and responds to any of the many sounds or "audio indications" in the form of a unique voices of a specific speaking attendees with the same command which is move the camera and highlight the PIP of the speaking attendee it reads on this broad language of "audio indication" or "audio command". In any case the rejection was based on two references whereby Pavlovic was clearly used to demonstrate the concept of a system utilizing a combination of both "audio commands" and a "related gesture" from a user as a means of controlling a graphical object on display which is analysis to where Inagaki controlled a specific graphical object such as a PIP on a display.

In response to applicant's argument on pages 11-13 that there is no suggestion to combine the references, the examiner recognizes that obviousness can only be established by combining or modifying the teachings of the prior art to produce the claimed invention where there is some teaching, suggestion, or motivation to do so found either in the references themselves or in the knowledge generally available to one of ordinary skill in the art. See *In re Fine*, 837 F.2d 1071, 5 USPQ2d 1596 (Fed. Cir. 1988) and *In re Jones*, 958 F.2d 347, 21 USPQ2d 1941 (Fed. Cir. 1992). In this case, Inagaki and Pavlovic are working in the same analysis field of art and are solving similar problems and therefore it is reasonable to think one of ordinary skill in the art working in this field would have had both references in front of them and would have been motivated from the references to combine features from each reference to come up with applicants proposed apparatus. And in this case examiner specifically found and

pointed out motivation from the secondary reference as to why one of ordinary skill in the art would have been motivated to improve the primary reference with features taught in the secondary reference. Note a suggestion that comes from one of the references used is considered the strongest source of motivation the examiner can use as the basis for the obvious combination.

In response to applicant's arguments on pages 13-14 that the examiner's conclusion of obviousness is based upon improper hindsight reasoning, it must be recognized that any judgment on obviousness is in a sense necessarily a reconstruction based upon hindsight reasoning. But so long as it takes into account only knowledge which was within the level of ordinary skill at the time the claimed invention was made, and does not include knowledge gleaned only from the applicant's disclosure, such a reconstruction is proper. See *In re McLaughlin*, 443 F.2d 1392, 170 USPQ 209 (CCPA 1971).

The applicant argues on pages 14-16, with regard to independent claim 11, that neither Inagaki or Pavlovic teach or suggest; "determining whether the received audio indication is one of a plurality of expected audio indications: analyzing a gesture of the user if the received audio indication is one of the plurality of expected audio indications" And also applicant argues that Independent claims 20 and 15 contain similar recitations regarding this "analyzing feature".

The examiner disagrees because Pavlovic figure 7 illustrates a plurality of "expected audio indications" SPEECH, and a plurality of "expected gestures" GESTURE. Now look at Pavlovic figure 5 where he illustrates in the audio estimator/

classifier module receiving and “determining whether the received audio indication is one of a plurality of expected audio indications” and where also he illustrates in the video estimator/ classifier module receiving and “determining whether the received gesture is one of a plurality of expected gestures” . It is an obvious practice that if either data collection process produces an error because the audio indication or gesture used is not from the expected sets illustrated in figure 7 that the next step of “analyzing a gesture of the user if the received audio indication is one of the plurality of expected audio” in the Feature Integrator will not happen. This is because it is an obvious practice when an artificial intelligent or smart device as illustrated by the combination of Inagaki/Pavlovic can not comprehend the data within a reasonable range of certainly or as stated by Pavlovic “computes the likelihood” that it simply errors out in the flow chart and does nothing but waits for further inputs.

The applicant argues on page 16, with regard to dependent claim 2, that neither Inagaki or Pavlovic teach or suggest, “the PIP display characteristic is at least one of a position of the PIP on the display and a display size of the PIP”. The examiner disagrees because the combination of Inagaki/Pavlovic clearly illustrates at least one of these changes, for example; “a position of the PIP on the display” figure 8a illustrates the concept of a relationship between who is speaking and the position of which PIP to be highlighted which further changes when speaker changes which is clearly illustrated in figure 8B so therefore it reads on this broad language.

The applicant argues on pages 16-17, with regard to dependent claim 5, that neither Inagaki or Pavlovic teach or suggest, “the processor is configured to analyze

image information received from the user after the audio indication is received to identify the change in the PIP display characteristic that is expressed by the received gesture.” With further regard to this feature applicant further argues that “Pavlovic teaches the user issuing a spoken command and gesture simultaneously”.

The examiner does not agree with applicant on this issue because how the data is collected for example simultaneously is irrelevant in fact applicant's specification also teaches he collects or receives his data simultaneously before he sends it for analyzing it is not clear the point of this argument in view of broadly written claim . Pavlovic figure 5, “HIGH LEVEL FEATURE INTEGRATION” where it was obvious the pre analyze step is to simultaneously receive the video and audio data using the camera and the microphone, where it is then later split into a parallel visual and audio estimator/classifier module which is followed by a second stage which contains a feature integration/combination module where the combination module computes the likelihood of the pairs of gesture and verbal words. This claim language is very broad here because Pavlovic clearly receives both the audio and video before he analyzes the video or audio data, this is just the logical progression claimed.

For the above reasons, it is believed that the rejections should be sustained.

Respectfully submitted,



Paul A. Bell  
Assistant Examiner  
October 29, 2004



Michael Razavi  
Supervisory Patent Examiner

Bipin Shalwala  
Conferee



BIPIN SHALWALA  
SUPERVISORY PATENT EXAMINER  
TECHNOLOGY CENTER 2600

PHILIPS INTELLECTUAL PROPERTY & STANDARDS  
P.O. BOX 3001  
BRIARCLIFF MANOR, NY 10510